

# BUBBLE STABILIZED DISCONTINUOUS GALERKIN METHOD FOR PARABOLIC AND ELLIPTIC PROBLEMS

ERIK BURMAN AND BENJAMIN STAMM

ABSTRACT. In this paper we give an analysis of a bubble stabilized discontinuous Galerkin method (BSDG) for elliptic and parabolic problems. The method consists of stabilizing the numerical scheme by enriching the discontinuous finite element space elementwise by quadratic non-conforming bubbles. This approach leads to optimal convergence in the space and time discretization parameters. Moreover the divergence of the diffusive fluxes converges in the  $L^2$ -norm independently of the geometry of the domain.

## 1. INTRODUCTION

The discontinuous Galerkin method (DG) for  $(2n)$ th-order elliptic problems was introduced by Baker [3] with special focus on the fourth order case. In parallel, the interior penalty method of Douglas and Dupont for second order elliptic problems [7] led to the symmetric interior penalty DG-method (SIPG) proposed by Wheeler [14] and Arnold [2]. In the interior penalty DG-method interelement continuity is enforced by adding a consistent stabilizing term on the interelement jump of the solution. To ensure coercivity the parameter in front of this penalty term must be large enough. On the other hand a large penalty term leads to a more ill-conditioned system.

In the nineties Oden et al. [9] introduced a non-symmetric DG-method (NIPG). The rationale for the non-symmetric version was that the finite element form was positive semi-definite without penalty term. Indeed it has been shown that the NIPG-method has optimal convergence in the broken energy-norm for polynomial orders  $p \geq 2$  [8, 12]. In a recent paper Brezzi and Marini [4] showed that the smallest space for which optimal convergence for the non-symmetric version is obtained is the space of piecewise affine functions enriched with quadratic bubbles. The stability properties in the case of piecewise affine approximation was given a precise characterization by Burman and Stamm [5]. In [5] we also proposed to enrich the piecewise affine space with the quadratic bubble, however we showed that for the enriched space the symmetric formulation is wellposed without penalty term and, thanks to the symmetry, optimally convergent in the  $L^2$ -norm. In this case on the other hand the matrix is not positive definite. The system matrix has negative eigenvalues, making the usefulness of the method questionable for problems with reaction terms or time-dependent problems. In principle a low order perturbation of the operator could shift the spectrum arbitrary close to zero leading to a singular, or nearly singular, matrix. The aim of the present paper is to show that

---

*Date:* April 8, 2008.

*Key words and phrases.* discontinuous Galerkin methods, parabolic and elliptic problems, bubble stabilization.

nevertheless, full stability without interior penalty term can be guaranteed under mild assumptions on the discretization parameters. We give a full analysis of the elliptic problem as well as the parabolic problem. For the special case of a right hand side that is piecewise constant per element we also show equivalence between the bubble stabilized symmetric formulation, the bubble stabilized non-symmetric formulation, a discontinuous Galerkin method with continuity imposed using Lagrange multipliers and finally the standard mixed finite element method using the lowest order Raviart-Thomas element for the fluxes and piecewise constant elements for the primal variable. All these different formulations are satisfied by the same finite element solution. This last result can be seen as the DG-version of work of Arbogast and Chen [1] in which they relate non-conforming finite element methods to mixed methods.

## 2. NOTATION

Let  $\Omega$  be a polygonal domain (polyhedron in three space dimensions) in  $\mathbb{R}^d$ ,  $d = 2, 3$ , with outer normal  $n$ . Let  $\mathcal{K}$  be a subdivision of  $\Omega \subset \mathbb{R}^d$  into non-overlapping  $d$ -simplices  $\kappa$  and denote by  $N_{\mathcal{K}}$  the number of simplices of the mesh. Suppose that each  $\kappa \in \mathcal{K}$  is an affine image of the reference element  $\hat{\kappa}$ , i.e. for each element  $\kappa$  there exists an affine transformation  $T_{\kappa} : \hat{\kappa} \rightarrow \kappa$ .

Let  $\mathcal{F}_i$  denote the set of interior faces ( $(d-1)$ -manifolds) of the mesh, i.e. the set of faces that are not included in the boundary  $\partial\Omega$ . The set  $\mathcal{F}_e$  denotes the faces that are included in  $\partial\Omega$  and define  $\mathcal{F} = \mathcal{F}_i \cup \mathcal{F}_e$ . Define by  $N_{\mathcal{F}} = \text{card}(\mathcal{F})$  and  $N_{\mathcal{F}_i} = \text{card}(\mathcal{F}_i)$  the number of faces resp. interior faces of the mesh. Denote by  $\Gamma$  the skeleton of the mesh, i.e. the set of points belonging to faces,  $\Gamma = \{x \in \overline{\Omega} \mid \exists F \in \mathcal{F} \text{ s.t. } x \in F\}$ .

Assume that  $\mathcal{K}$  is shape-regular, does not contain any hanging node and covers  $\overline{\Omega}$  exactly. For an element  $\kappa \in \mathcal{K}$ ,  $h_{\kappa}$  denotes its diameter and for a face  $F \in \mathcal{F}$ ,  $h_F$  denotes the diameter of  $F$ . Set  $h = \max_{\kappa \in \mathcal{K}} h_{\kappa}$  and let  $\tilde{h}$  be the function such that  $\tilde{h}|_{\kappa} = h_{\kappa}$  and  $\tilde{h}|_F = h_F$  for all  $\kappa \in \mathcal{K}$  and  $F \in \mathcal{F}$ .

For a subset  $R \subset \Omega$  or  $R \subset \mathcal{F}$ ,  $(\cdot, \cdot)_R$  denotes the  $L^2(R)$ -scalar product,  $\|\cdot\|_R = (\cdot, \cdot)_R^{1/2}$  the corresponding norm, and  $\|\cdot\|_{s,R}$  the  $H^s(R)$ -norm. The element-wise counterparts will be distinguished using the discrete partition as subscript, for example  $(\cdot, \cdot)_{\mathcal{K}} = \sum_{\kappa \in \mathcal{K}} (\cdot, \cdot)_{\kappa}$ . For  $s \geq 1$ , let  $H^s(\mathcal{K})$  be the space of piecewise Sobolev  $H^s$ -functions and denote its norm by  $\|\cdot\|_{s,\mathcal{K}}$ .

In this paper  $c > 0$  denotes a generic constant and can change at each occurrence, while an indexed constant stays fix. Any constant is independent of the mesh size  $h$ .

Further let us define the jump and average operators. Fix  $F \in \mathcal{F}_i$  and thus  $F = \kappa_1 \cap \kappa_2$  with  $\kappa_1, \kappa_2 \in \mathcal{K}$ . Let  $v \in H^1(\mathcal{K})$  and denote by  $v_1, v_2$  the restriction of  $v$  to the element  $\kappa_1, \kappa_2$ , i.e.  $v_1 = v|_{\kappa_1}$  resp.  $v_2 = v|_{\kappa_2}$  and denote by  $n_1, n_2$  the exterior normal of  $\kappa_1$  resp.  $\kappa_2$ . Then we define the standard average and jump operators by

$$\begin{aligned} \{v\} &= \frac{1}{2}(v_1 + v_2), & [v] &= v_1 n_1 + v_2 n_2, \\ \{\nabla v\} &= \frac{1}{2}(\nabla v_1 + \nabla v_2), & [\nabla v] &= \nabla v_1 \cdot n_1 + \nabla v_2 \cdot n_2. \end{aligned}$$

On outer faces  $F \in \mathcal{F}_e$  we define them by

$$\{v\} = v, \quad [v] = vn, \quad \{\nabla v\} = \nabla v, \quad [\nabla v] = \nabla v \cdot n$$

where  $n$  is the outer normal of the domain  $\Omega$ . Observe that  $[v](x), \{\nabla v\}(x) \in \mathbb{R}^d$ . Thus we introduce also scalar quantities of the jump and average, that we index by  $s$ :

$$[v]_s = [v] \cdot n_F, \quad \{\nabla v\}_s = \{\nabla v\} \cdot n_F,$$

where  $n_F \in \{n_1, n_2\}$  is arbitrarily chosen but fixed. Observe that also  $[v] = [v]_s n_F$  and thus that  $\|[v]\|_{\mathcal{F}} = \|[v]_s\|_{\mathcal{F}}$  and  $\|[v]\|_{\mathcal{F}_i} = \|[v]_s\|_{\mathcal{F}_i}$ . Moreover note that

$$(1) \quad [v] \cdot \{\nabla w\} = [v]_s \{\nabla w\}_s$$

for all  $v, w \in H^1(\mathcal{K})$ . The following integration by part holds.

**Lemma 2.1** (Integration by parts). *Let  $v, w \in H^k(\mathcal{K})$ , then*

$$(2) \quad (\nabla v, \nabla w)_{\mathcal{K}} = -(\Delta v, w)_{\mathcal{K}} + (\{\nabla v\}, [w])_{\mathcal{F}} + ([\nabla v], \{w\})_{\mathcal{F}_i}.$$

*Proof.* Equality (2) results from element-wise integration by parts and applying the definitions of the standard jump and average operators.  $\square$   $\square$

### 3. BUBBLE STABILIZED FINITE ELEMENT SPACE

Let us denote by  $V_h^p$  the standard discontinuous finite element space of degree  $p \geq 0$  defined by

$$V_h^p = \{v_h \in L^2(\Omega) \mid v_h|_{\kappa} \in \mathbb{P}_p(\kappa), \forall \kappa \in \mathcal{K}\},$$

where  $\mathbb{P}_p(\kappa)$  denotes the set of polynomials of maximum degree  $p$  on  $\kappa$ . Consider the enriched finite element space

$$V_{bs} = V_h^1 \oplus \{v_h \in L^2(\Omega) \mid v_h(x) = \alpha x \cdot x, \alpha \in V_h^0\},$$

where  $x = (x_1, \dots, x_d)$  denotes the physical variable. Let us additionally define some functional space that consists of functions only defined on the skeleton of the mesh:

$$W_h^0 = \{v_h \in L^2(\Gamma) \mid v_h|_F \in \mathbb{P}_0(F), \forall F \in \mathcal{F}\}.$$

Let  $v \in H^1(\mathcal{K})$  and define by  $\overline{\{v\}}, \overline{[v]}_s$  the  $L^2$ -projection of  $\{v\}$  resp.  $[v]_s$  onto  $W_h^0$ , i.e.

$$\begin{aligned} \overline{\{v\}}, w_h)_{\mathcal{F}} &= (\{v\}, w_h)_{\mathcal{F}}, & \forall w_h \in W_h^0, \\ (\overline{[v]}_s, w_h)_{\mathcal{F}} &= ([v]_s, w_h)_{\mathcal{F}}, & \forall w_h \in W_h^0. \end{aligned}$$

Further define  $\overline{[v]} = \overline{[v]}_s n_F$ .

We denote by  $RT_0(\kappa)$  the local Raviart-Thomas space of order 0 defined by

$$RT_0(\kappa) = [\mathbb{P}_0(\kappa)]^d + x \mathbb{P}_0(\kappa)$$

and its global variant as well as the Crouzeix-Raviart space by

$$\begin{aligned} RT_0 &= \{v \in [L^1(\Omega)]^d \mid v|_{\kappa} \in RT_0(\kappa) \forall \kappa \in \mathcal{K}, [v]|_F = 0, \forall F \in \mathcal{F}_i\}, \\ CR &= \{v \in L^1(\Omega) \mid v|_{\kappa} \in \mathbb{P}_1(\kappa) \forall \kappa \in \mathcal{K}, \overline{[v]}|_F = 0, \forall F \in \mathcal{F}_i\}. \end{aligned}$$

Then we denote by  $H^s(\Omega) + CR$ ,  $s \in \{1, 2\}$ , the space of functions  $v$  such that there exists a (non-unique) decomposition  $v = v_1 + v_2$  with  $v_1 \in H^s(\Omega)$  and  $v_2 \in CR$ .

**3.1. Properties of bubble stabilized finite element space.** Let us discuss some important properties of the space  $V_{bs}$ .

**Lemma 3.1.** *For  $v_h \in V_{bs}$  we have that*

$$\Delta v_h \in V_h^0.$$

*Moreover the application  $\Delta : V_{bs} \rightarrow V_h^0$  is surjective.*

*Proof.* Observe that  $\Delta w_h = 0$  for all  $w_h \in V_h^1$  and that  $\Delta(\alpha x \cdot x) = 2d\alpha \in V_h^0$  where  $d$  is the dimension of  $\Omega$ .  $\square$

**Lemma 3.2.** *For all  $v_h \in V_{bs}$  there holds*

$$\nabla v_h|_\kappa \in RT_0(\kappa), \quad \forall \kappa \in \mathcal{K},$$

*and for all  $\kappa \in \mathcal{K}$  and  $r_h \in RT_0(\kappa)$ , there exists  $v_h \in V_{bs}$  such that  $\nabla v_h|_\kappa = r_h$ .*

*Proof.* We refer to the proof of Lemma 3.4 in [5].  $\square$

**Corollary 3.3.** *For  $v_h \in V_{bs}$  we have that*

$$\{\nabla v_h\}_s \in W_h^0, \quad [\nabla v_h] \in W_h^0.$$

*Moreover the applications  $\{\nabla \cdot\}_s : V_{bs} \rightarrow W_h^0$  and  $[\nabla \cdot] : V_{bs} \rightarrow W_h^0$  are surjective.*

**Lemma 3.4.** *There is a constant  $c > 0$  independent of  $h$  such that for all  $v_h \in V_{bs}$  there holds*

$$\|\tilde{h}^{-\frac{1}{2}}[v_h]\|_{\mathcal{F}}^2 \leq c \left( \|\nabla v_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}}\overline{[v_h]}\|_{\mathcal{F}}^2 \right).$$

*Proof.* We refer to the proof of Lemma 4.1 in [5].  $\square$

**Lemma 3.5** (Poincaré inequality). *There is a constant  $c_P > 0$  independent of  $h$  such that for all  $v_h \in V_{bs}$  there holds*

$$\|v_h\|_{\mathcal{K}}^2 \leq c_P \left( \|\nabla v_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}}\overline{[v_h]}\|_{\mathcal{F}}^2 \right).$$

*Proof.* We refer to the proof of Corollary 4.2 in [5].  $\square$

Observe that Lemma 3.4 and 3.5 are only valid on discrete spaces and thus does not hold for functions in  $H^1(\mathcal{K})$ . Thus we define the following norm on  $H^1(\mathcal{K})$  by

$$|||v|||^2 = \|\nabla v\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}}[v]\|_{\mathcal{F}}^2$$

for all  $v \in H^1(\mathcal{K})$ . Nevertheless observe that there exists a constant  $c_d > 0$  such that

$$(3) \quad c_d |||v_h|||^2 \leq \|\nabla v_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}}\overline{[v_h]}\|_{\mathcal{F}}^2 \leq |||v_h|||^2$$

for all  $v_h \in V_{bs}$ .

**3.2. Technical lemmas.** In this section we recall some well known results. For the proofs we refer to [6].

**Lemma 3.6** (Inverse inequality). *Let  $v_h \in V_{bs}$ , then there exists a constant  $c_I > 0$  independent of  $h$  such that*

$$c_I^{-1} \|\tilde{h}^2 \Delta v_h\|_{\mathcal{K}} \leq \|\tilde{h} \nabla v_h\|_{\mathcal{K}} \leq c_I \|v_h\|_{\mathcal{K}}.$$

**Lemma 3.7** (Trace inequality). *Let  $v \in [H^1(\mathcal{K})]^m$  and  $v_h \in [V_{bs}]^m$  with  $m \geq 1$ , then there exists a constant  $c_T > 0$  independent of  $h$  such that*

$$\begin{aligned} \|\{v\}\|_{\mathcal{F}} + \|[v]\|_{\mathcal{F}} &\leq c_T \left( \|\tilde{h}^{-\frac{1}{2}} v\|_{\mathcal{K}} + \|\tilde{h}^{\frac{1}{2}} \nabla v\|_{\mathcal{K}} \right), \\ \|\{v_h\}\|_{\mathcal{F}} + \|[v_h]\|_{\mathcal{F}} &\leq c_T \|\tilde{h}^{-\frac{1}{2}} v_h\|_{\mathcal{K}}. \end{aligned}$$

**3.3. Projections.** We denote by  $\pi_p : L^2(\Omega) \rightarrow V_h^p$  the  $L^2$ -projection onto  $V_h^p$  defined by

$$\int_{\Omega} \pi_p(v) w_h \, dx = \int_{\Omega} v w_h \, dx \quad \forall w_h \in V_h^p.$$

Then  $\pi_p$  satisfies the following approximation result: Let  $v \in H^{p+1}(\mathcal{K})$ , then

$$(4) \quad \|v - \pi_p v\|_{\mathcal{K}} + h \|\nabla(v - \pi_p v)\|_{\mathcal{K}} \leq ch^{p+1} |v|_{p+1, \mathcal{K}}.$$

Additionally let us denote by  $i_c : H^1(\Omega) \rightarrow CR$  the Crouzeix-Raviart interpolant interpolating on the midpoints of each internal face satisfying the following approximation result: let  $v \in H^2(\Omega)$ , then

$$(5) \quad \|v - i_c v\|_{\mathcal{K}} + h \|\nabla(v - i_c v)\|_{\mathcal{K}} \leq ch^2 |v|_{2, \mathcal{K}}.$$

Finally, we present two projections which will be used in the following.

**Lemma 3.8.** *Let  $a_h \in V_h^0$  and  $b_h, c_h \in W_h^0$  be fixed. Then, there exists a unique function  $\phi_h \in V_{bs}$  such that*

$$(6) \quad \begin{cases} \pi_0 \phi_h &= a_h, \\ \overline{[\phi_h]_s}|_F &= b_h|_F \quad \forall F \in \mathcal{F}, \\ \{\phi_h\}|_F &= c_h|_F \quad \forall F \in \mathcal{F}_i. \end{cases}$$

Moreover  $\phi_h$  satisfies the following stability result

$$(7) \quad \|\tilde{h}^{-1} \phi_h\|_{\mathcal{K}}^2 + \|[\phi_h]\|_{\mathcal{F}}^2 \leq c \left( \|\tilde{h}^{-1} a_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}} b_h\|_{\mathcal{F}}^2 + \|\tilde{h}^{-\frac{1}{2}} c_h\|_{\mathcal{F}_i}^2 \right).$$

*Proof.* Let us first establish the a priori estimate. Observe that

$$(8) \quad \|\tilde{h}^{-1} \phi_h\|_{\mathcal{K}}^2 \leq \|\tilde{h}^{-1} \pi_0 \phi_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-1} (\phi_h - \pi_0 \phi_h)\|_{\mathcal{K}}^2 \leq \|\tilde{h}^{-1} a_h\|_{\mathcal{K}}^2 + c \|\nabla \phi_h\|_{\mathcal{K}}^2.$$

By equation (3), integration by parts, Lemma 3.1, equation (1) and Corollary 3.3 it follows that

$$\begin{aligned} c_d \|[\phi_h]\|_{\mathcal{F}}^2 &\leq \|\tilde{h}^{-\frac{1}{2}} \overline{[\phi_h]}\|_{\mathcal{F}}^2 + \|\nabla \phi_h\|_{\mathcal{K}}^2 \\ &= \|\tilde{h}^{-\frac{1}{2}} b_h\|_{\mathcal{F}}^2 - (\Delta \phi_h, \pi_0 \phi_h)_{\mathcal{K}} + (\{\nabla \phi_h\}_s, \overline{[\phi_h]_s})_{\mathcal{F}} + ([\nabla \phi_h], \overline{\{\phi_h\}})_{\mathcal{F}_i} \\ &= \underbrace{\|\tilde{h}^{-\frac{1}{2}} b_h\|_{\mathcal{F}}^2 - (\Delta \phi_h, a_h)_{\mathcal{K}}}_I + \underbrace{(\{\nabla \phi_h\}_s, b_h)_{\mathcal{F}}}_{II} + \underbrace{([\nabla \phi_h], c_h)_{\mathcal{F}_i}}_{III}. \end{aligned}$$

Applying the Cauchy-Schwarz, the inverse (I) or the trace (II, III) and then Young's inequality for each term yields respectively

$$\begin{aligned} I &\leq c_I \|\nabla \phi_h\|_{\mathcal{K}} \|\tilde{h}^{-1} a_h\|_{\mathcal{K}} \leq \frac{1}{4} \|\nabla \phi_h\|_{\mathcal{K}}^2 + c_I^2 \|\tilde{h}^{-1} a_h\|_{\mathcal{K}}^2 \\ II &\leq c_T \|\nabla \phi_h\|_{\mathcal{K}} \|\tilde{h}^{-\frac{1}{2}} b_h\|_{\mathcal{F}} \leq \frac{1}{4} \|\nabla \phi_h\|_{\mathcal{K}}^2 + c_T^2 \|\tilde{h}^{-\frac{1}{2}} b_h\|_{\mathcal{F}}^2, \\ III &\leq c_T \|\nabla \phi_h\|_{\mathcal{K}} \|\tilde{h}^{-\frac{1}{2}} c_h\|_{\mathcal{F}_i} \leq \frac{1}{4} \|\nabla \phi_h\|_{\mathcal{K}}^2 + c_T^2 \|\tilde{h}^{-\frac{1}{2}} c_h\|_{\mathcal{F}_i}^2, \end{aligned}$$

and thus, combining with (8), yields

$$\|\tilde{h}^{-1} \phi_h\|_{\mathcal{K}}^2 + \|\phi_h\|^2 \leq c \left( \|\tilde{h}^{-1} a_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}} b_h\|_{\mathcal{F}}^2 + \|\tilde{h}^{-\frac{1}{2}} c_h\|_{\mathcal{F}_i}^2 \right).$$

To conclude the proof, it now suffices to observe that (6) is nothing more than a square linear system of size  $N_{\mathcal{K}} + N_{\mathcal{F}} + N_{\mathcal{F}_i}$ . Hence, existence and uniqueness of a solution of the linear system are equivalent. Let us denote by  $Aw = b$  the square linear system and assume that there is a vector  $w_1$  and  $w_2$  such that  $Aw_i = b$ ,  $i = 1, 2$ . Further let us denote the difference between them by  $e = w_1 - w_2$  and therefore  $Ae = 0$ . The a priori estimate (7) implies that  $e = 0$  and thus the solution is unique and hence the matrix is regular.  $\square$   $\square$

Using a similar technique one can also prove the existence of the following projection.

**Lemma 3.9.** *Let  $a_h \in V_h^0$  and  $b_h, c_h \in W_h^0$  be fixed. Then, there exists a unique function  $\phi_h \in V_{bs}$  such that*

$$(9) \quad \begin{cases} \pi_0 \phi_h &= a_h, \\ \{\nabla \phi_h\}_s|_F &= b_h|_F \quad \forall F \in \mathcal{F}, \\ \{\phi_h\}|_F &= c_h|_F \quad \forall F \in \mathcal{F}_i. \end{cases}$$

Moreover  $\phi_h$  satisfies the following stability result

$$(10) \quad \|\tilde{h}^{-1} \phi_h\|_{\mathcal{K}}^2 + \|\phi_h\|^2 \leq c \left( \|\tilde{h}^{-1} a_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{\frac{1}{2}} b_h\|_{\mathcal{F}}^2 + \|\tilde{h}^{-\frac{1}{2}} c_h\|_{\mathcal{F}_i}^2 \right).$$

*Proof.* The proof in this case is similar as the one of Lemma 3.8. Details are left to the reader.  $\square$   $\square$

#### 4. ELLIPTIC PROBLEMS

Let us consider the following diffusion-reaction equation: find  $u : \Omega \rightarrow \mathbb{R}$  such that

$$(11) \quad \begin{cases} -\Delta u + \tau u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{cases}$$

where  $\tau \geq 0$  and  $f \in L^2(\Omega)$ . Assuming a homogenous boundary condition and a diffusion coefficient of one is not a restriction but simplifies the presentation of the method.

**4.1. Bubble stabilized discontinuous Galerkin method.** Consider the symmetric scheme: find  $u_h \in V_{bs}$  such that

$$(12) \quad A(u_h, v_h) = (f, v_h)_{\mathcal{K}} \quad \forall v_h \in V_{bs},$$

with

$$(13) \quad A(v, w) = a(v, w) + (\tau v, w)_{\mathcal{K}},$$

$$(14) \quad a(v, w) = (\nabla v, \nabla w)_{\mathcal{K}} - (\{\nabla v\}, [w])_{\mathcal{F}} - ([v], \{\nabla w\})_{\mathcal{F}},$$

for all  $v, w \in H^2(\mathcal{K})$ .

*Remark 4.1.* The discrete solution  $u_h$  of (12) satisfies the following local mass conservation property

$$\int_{\kappa} \tau u_h dx - \int_{\partial\kappa} \{\nabla u_h\} \cdot n_{\kappa} ds = \int_{\kappa} f dx \quad \forall \kappa \in \mathcal{K}.$$

**Lemma 4.2** (Consistency). *Let  $u \in H^2(\Omega)$  be the exact solution of (11) and let  $u_h \in V_{bs}$  be the approximation defined by (12), then the method is consistent and adjoint consistent. Moreover the following Galerkin orthogonality holds*

$$A(u - u_h, v_h) = 0 \quad \forall v_h \in V_{bs}.$$

*Proof.* We only show the primal consistency, dual consistency follows by symmetry. Since  $u \in H^2(\Omega)$  it follows that  $u \in [u]|_F = 0$  for all  $F \in \mathcal{F}$  and  $[\nabla u]|_F = 0$  for all  $F \in \mathcal{F}_i$ . Therefore by definition of the bilinear form and integration by parts

$$\begin{aligned} A(u, v_h) &= (\nabla u, \nabla v_h)_{\mathcal{K}} + (\tau u, v_h)_{\mathcal{K}} - (\{\nabla u\}, [v_h])_{\mathcal{F}} - ([u], \{\nabla v_h\})_{\mathcal{F}} \\ &= (-\Delta u + \tau u, v_h)_{\mathcal{K}} + ([\nabla u], \{v_h\})_{\mathcal{F}_i} \\ &= (f, v_h)_{\mathcal{K}} = A(u_h, v_h). \end{aligned}$$

□

□

**4.2. Stability.** Associated to the bilinear form  $A(\cdot, \cdot)$  define the following energy-norm:

$$(15) \quad |||v|||_e^2 = \frac{c_d}{4} |||v|||^2 + \|\tau^{\frac{1}{2}} v\|_{\mathcal{K}}^2$$

for all  $v \in H^1(\mathcal{K})$  and where  $c_d > 0$  denotes the constant of (3). Furthermore, define the following norm

$$|||v|||_c^2 = \|\tilde{h}^{\frac{1}{2}} \{\nabla v\}\|_{\mathcal{F}}^2 + |||v|||_e^2$$

for all  $v \in H^1(\mathcal{K})$ . As a consequence of the trace inequality and (5) there holds

$$(16) \quad |||v - i_c v|||_c \leq ch|v|_{2,\mathcal{K}}.$$

**Proposition 4.3** (Inf-sup condition). *There exists a constant  $c > 0$  independent of  $h$  such that for  $\tau = 0$  or,  $\tau > 0$  with  $h^2 < c_s/\tau$  for some constant  $c_s > 0$  independent of  $h$  and  $\tau$ , there holds*

$$\forall v_h \in V_{bs}, \quad c |||v_h|||_e \leq \sup_{0 \neq w_h \in V_{bs}} \frac{A(v_h, w_h)}{|||w_h|||_e}.$$

*Proof.* For the proof of Proposition 4.3 we introduce the following two lemmas:

**Lemma 4.4.** *For  $\tau = 0$  or,  $\tau > 0$  with  $h^2 < c_s/\tau$  for some constant  $c_s > 0$  independent of  $h$  and  $\tau$ , there exists for each fixed  $v_h \in V_{bs}$  a function  $w_h \in V_{bs}$  such that*

$$|||v_h|||_e^2 \leq A(v_h, w_h).$$

**Lemma 4.5.** *Let  $v_h$  and  $w_h$  be the functions defined in Lemma 4.4, then there exists a constant  $c > 0$  independent of  $h$  such that*

$$|||w_h|||_e \leq c |||v_h|||_e.$$

Combining Lemma 4.4 and 4.5 leads to the result. Indeed, for a fixed  $v_h$  there holds

$$A(v_h, w_h) \geq \|v_h\|_e^2 \geq c \|v_h\|_e \|w_h\|_e.$$

□

□

*Proof of Lemma 4.4.* Let  $v_h \in V_{bs}$  be fixed. Then, by the Cauchy-Schwarz, Young's and the trace inequality we get

$$\begin{aligned} A(v_h, v_h) &= \|\nabla v_h\|_{\mathcal{K}}^2 + \|\tau^{\frac{1}{2}} v_h\|_{\mathcal{K}}^2 - 2(\{\nabla v_h\}, \overline{[v_h]})_{\mathcal{F}} \\ (17) \quad &\geq \frac{1}{2} \|\nabla v_h\|_{\mathcal{K}}^2 + \|\tau^{\frac{1}{2}} v_h\|_{\mathcal{K}}^2 - 2c_T^2 \|\tilde{h}^{-\frac{1}{2}} \overline{[v_h]}\|_{\mathcal{F}}^2 \end{aligned}$$

where  $c_T > 0$  is the constant from the trace inequality.

Secondly denote by  $\phi_h \in V_{bs}$  the projection defined by Lemma 3.9 with  $a_h = 0$ ,  $b_h = -\tilde{h}^{-1} \overline{[v_h]_s}$ ,  $c_h = 0$ . Then, by (1), integration by parts and Corollary 3.3

$$\begin{aligned} A(v_h, \phi_h) &= (\nabla v_h, \nabla \phi_h)_{\mathcal{K}} + (\tau v_h, \phi_h)_{\mathcal{K}} - (\{\nabla v_h\}, [\phi_h])_{\mathcal{F}} - ([v_h], \{\nabla \phi_h\})_{\mathcal{F}} \\ &= -(\Delta v_h, \pi_0 \phi_h)_{\mathcal{K}} + (\tau v_h, \phi_h)_{\mathcal{K}} + ([\nabla v_h], \overline{[\phi_h]})_{\mathcal{F}_i} - (\overline{[v_h]_s}, \{\nabla \phi_h\}_s)_{\mathcal{F}} \\ (18) \quad &= (\tau v_h, \phi_h)_{\mathcal{K}} + \|\tilde{h}^{-\frac{1}{2}} \overline{[v_h]}\|_{\mathcal{F}}^2. \end{aligned}$$

Combining (17) and (18) it follows that for  $\tau = 0$

$$A(v_h, w_h) \geq \frac{1}{2} (\|\nabla v_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}} \overline{[v_h]}\|_{\mathcal{F}}^2) \geq \frac{c_d}{2} \|v_h\|^2 \geq \|v_h\|_e^2,$$

with  $w_h = v_h + (2c_T^2 + \frac{1}{2})\phi_h$ . On the other hand if  $\tau > 0$  we have

$$(\tau v_h, w_h)_{\mathcal{K}} = \|\tau^{\frac{1}{2}} v_h\|_{\mathcal{K}}^2 + c(\tau v_h, \phi_h)_{\mathcal{K}}.$$

Since  $\pi_0 \phi_h = 0$  and by the stability estimate (10) yields

$$\begin{aligned} (\tau v_h, \phi_h)_{\mathcal{K}} &= \tau(v_h - \pi_0 v_h, \phi_h)_{\mathcal{K}} \leq \tau \|v_h - \pi_0 v_h\|_{\mathcal{K}} \|\phi_h\|_{\mathcal{K}} \\ &\leq c\tau h^2 \|\nabla v_h\|_{\mathcal{K}} \|\tilde{h}^{-\frac{1}{2}} \overline{[v_h]}\|_{\mathcal{F}} \leq c\tau h^2 (\|\nabla v_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}} \overline{[v_h]}\|_{\mathcal{F}}^2). \end{aligned}$$

Thus

$$\begin{aligned} A(v_h, w_h) &\geq (\frac{1}{2} - c\tau h^2) (\|\nabla v_h\|_{\mathcal{K}}^2 + \|\tilde{h}^{-\frac{1}{2}} \overline{[v_h]}\|_{\mathcal{F}}^2) + \|\tau^{\frac{1}{2}} v_h\|_{\mathcal{K}}^2 \\ &\geq c_d (\frac{1}{2} - c\tau h^2) \|v_h\|^2 + \|\tau^{\frac{1}{2}} v_h\|_{\mathcal{K}}^2. \end{aligned}$$

Therefore under assumption  $h < \frac{1}{\sqrt{4c\tau}}$  we may conclude

$$A(v_h, w_h) \geq \|v_h\|_e^2.$$

□

□

*Proof of Lemma 4.5.* Using the triangle inequality and the stability estimate (10) observe that

$$\|w_h\|_e \leq \|v_h\|_e + c \|\phi_h\|_e \leq c \|v_h\|_e$$

since  $h^2\tau \leq c$ .

□

□

**Corollary 4.6.** *Under the assumption of Proposition 4.3, the discrete problem (12) has a unique solution. Furthermore the following estimation holds*

$$\|u_h\|_e \leq c \|f\|_{\mathcal{K}}$$

where  $c > 0$  is a constant independent of  $h$ .



*Proof.* Observe that by the inf-sup condition, the Cauchy-Schwarz inequality and the Poincaré inequality

$$\begin{aligned} |||u_h|||_e &\leq c \sup_{0 \neq w_h \in V_{bs}} \frac{A(u_h, w_h)}{|||w_h|||_e} \leq c \sup_{0 \neq w_h \in V_{bs}} \frac{(f, w_h)_K}{|||w_h|||_e} \leq c \|f\|_K \sup_{0 \neq w_h \in V_{bs}} \frac{\|w_h\|_K}{|||w_h|||_e} \\ &\leq c \|f\|_K. \end{aligned}$$

□

□

#### 4.3. Convergence analysis.

**Lemma 4.7** (Continuity). *For  $v, w \in H^1(K)$  there holds*

$$|A(v, w)| \leq c |||v|||_c |||w|||_c.$$

*Further in the particular case of  $v \in H^1(\Omega) + CR$  and  $w_h \in V_{bs}$  there holds*

$$|A(v, w_h)| \leq c |||v|||_c |||w_h|||_e.$$

*Proof.* This is an immediate consequence of the definition of the bilinear form  $A(\cdot, \cdot)$  and the Cauchy-Schwarz inequality which leads to

$$|A(v, w)| = |(\nabla v, \nabla w)_K + (\tau v, w)_K - (\{\nabla v\}, [w])_{\mathcal{F}} - (\{\nabla w\}, [v])_{\mathcal{F}}| \leq c |||v|||_c |||w|||_c$$

and

$$|A(v, w_h)| = |(\nabla v, \nabla w_h)_K + (\tau v, w_h)_K - (\{\nabla v\}, [w_h])_{\mathcal{F}}| \leq c |||v|||_c |||w_h|||_e.$$

□

□

**Theorem 4.8** (Convergence). *Let  $u \in H^2(\Omega)$  be the solution of (11) and  $u_h$  be the discrete solution of (12). For  $\tau = 0$  or,  $\tau > 0$  with  $h^2 < c_s/\tau$  for some constant  $c_s > 0$  independent of  $h$  and  $\tau$ , there holds*

$$|||u - u_h|||_e \leq c h |u|_{2,K}.$$

*Proof.* Let us split the error in two parts

$$u - u_h = (u - i_c u) + (i_c u - u_h) = \eta + \xi_h,$$

with  $\eta = u - i_c u \in H^2(\Omega) + CR$  and  $\xi_h = i_c u - u_h \in V_{bs}$ , and therefore

$$(19) \quad |||u - u_h|||_e \leq |||\eta|||_e + |||\xi_h|||_e \leq |||\eta|||_c + |||\xi_h|||_e.$$

Let us focus on the second term of the right hand side of (19). Using the inf-sup condition, the consistency and the continuity we may write

$$(20) \quad |||\xi_h|||_e \leq c \sup_{0 \neq w_h \in V_{bs}} \frac{A(\xi_h, w_h)}{|||w_h|||_e} = c \sup_{0 \neq w_h \in V_{bs}} \frac{|A(\eta, w_h)|}{|||w_h|||_e} \leq c |||\eta|||_c.$$

Thus by inserting (20) into (19) we get

$$|||u - u_h|||_e \leq c |||\eta|||_c \leq c h |u|_{2,K}.$$

using the approximation result (16). □

□

**Theorem 4.9** ( $L^2$ -convergence). *Let  $u \in H^2(\Omega)$  with  $|u|_{2,K} \leq c \|f\|_K$  be the solution of (11) and  $u_h$  the solution of (12). For  $\tau = 0$  or,  $\tau > 0$  with  $h^2 < c_s/\tau$  for some constant  $c_s > 0$  independent of  $h$  and  $\tau$ , there holds*

$$\|u - u_h\|_K \leq c h^2 |u|_{2,K}.$$

*Proof.* Let  $e = u - u_h$  and consider the dual problem: find  $\phi \in H_0^1(\Omega)$  such that

$$(\nabla \phi, \nabla z)_K + (\tau \phi, z)_K = (e, z)_K, \quad \forall z \in H_0^1(\Omega).$$

We note that  $\phi \in H^2(\Omega)$  and that by integration by parts there holds

$$-(\Delta \phi, z)_K + (\tau \phi, z)_K = (e, z)_K, \quad \forall z \in L^2(\Omega).$$

We may then choose  $z = e \in H^2(K)$  and after another integration by parts we write  $A(\phi, e) = \|e\|_K^2$  using the dual consistency of Lemma 4.2 and the regularity of  $e$  and  $\phi$ . By the regularity assumption on  $u$  it follows that  $|\phi|_{2,K} \leq c \|e\|_K$  and thus by the continuity of Lemma 4.7, the trace inequality, the approximability of the interpolation  $i_c$  and the a priori estimate of Theorem 4.8 we get

$$\begin{aligned} \|e\|_K^2 &= A(\phi, e) = A(\phi - i_c \phi, e) \leq c \| \phi - i_c \phi \|_c \|e\|_c \\ &\leq c (\| \phi - i_c \phi \|_e^2 + h^2 |\phi|_{2,K}^2)^{\frac{1}{2}} (\|e\|_e^2 + h^2 |u|_{2,K}^2)^{\frac{1}{2}} \leq c h^2 |\phi|_{2,K} |u|_{2,K} \\ &\leq c h^2 \|e\|_K |u|_{2,K}. \end{aligned}$$

□

□

**Proposition 4.10.** *Let  $u_h$  be the solution of (12). Then, the following estimation holds*

$$(21) \quad h \|f + \Delta u_h - \tau u_h\|_K + \|\tilde{h}^{\frac{1}{2}} [\nabla u_h]\|_{\mathcal{F}_i} + \|\tilde{h}^{-\frac{1}{2}} \overline{[u_h]}\|_{\mathcal{F}} \leq ch (\|f - \pi_0 f\|_K + \tau h \|f\|_K),$$

and if  $f \in H^1(K)$  there holds

$$(22) \quad h \|f + \Delta u_h - \tau u_h\|_K + \|\tilde{h}^{\frac{1}{2}} [\nabla u_h]\|_{\mathcal{F}_i} + \|\tilde{h}^{-\frac{1}{2}} \overline{[u_h]}\|_{\mathcal{F}} \leq ch^2 (\|\nabla f\|_K + \tau \|f\|_K).$$

*Proof.* Let  $\phi_h \in V_{bs}$  be the projection defined by Lemma 3.9 with  $a_h = h(\tau \pi_0 u_h - \pi_0 f - \Delta u_h)$ ,  $b_h = -\tilde{h}^{-1} \overline{[u_h]_s}$  and  $c_h = \tilde{h} [\nabla u_h]$ . Then, by integration by parts it follows that

$$\begin{aligned} A(u_h, \phi_h) &= -(\Delta u_h, \pi_0 \phi_h)_K + ([\nabla u_h], \overline{\{\phi_h\}})_{\mathcal{F}_i} - (\overline{[u_h]_s}, \{\nabla \phi_h\}_s)_{\mathcal{F}} + \tau(u_h, \phi_h)_K \\ &= (\tau \pi_0 u_h - \Delta u_h, a_h)_K + \|\tilde{h}^{\frac{1}{2}} [\nabla u_h]\|_{\mathcal{F}_i}^2 + \|\tilde{h}^{-\frac{1}{2}} \overline{[u_h]}\|_{\mathcal{F}}^2 + \tau(u_h - \tau \pi_0 u_h, \phi_h)_K, \end{aligned}$$

and thus since  $A(u_h, \phi_h) = (f, \phi_h)_K$  we get

$$\begin{aligned} h^2 \|\pi_0 f + \Delta u_h - \tau \pi_0 u_h\|_K^2 + \|\tilde{h}^{\frac{1}{2}} [\nabla u_h]\|_{\mathcal{F}_i}^2 + \|\tilde{h}^{-\frac{1}{2}} \overline{[u_h]}\|_{\mathcal{F}}^2 \\ = (f - \pi_0 f, \phi_h)_K - \tau(u_h - \pi_0 u_h, \phi_h)_K \\ \leq c (\|f - \pi_0 f\|_K + \tau \|u_h - \pi_0 u_h\|_K) \|\phi_h\|_K. \end{aligned}$$

Now observe firstly that

$$(23) \quad \|u_h - \pi_0 u_h\|_K \leq ch \|\nabla u_h\|_K \leq ch \|f\|_K$$

by Corollary 4.6 and secondly by the stability estimate (10) of  $\phi_h$  that

$$\|\phi_h\|_K \leq ch \left( h^2 \|\pi_0 f + \Delta u_h - \tau \pi_0 u_h\|_K^2 + \|\tilde{h}^{\frac{1}{2}} [\nabla u_h]\|_{\mathcal{F}_i}^2 + \|\tilde{h}^{-\frac{1}{2}} \overline{[u_h]}\|_{\mathcal{F}}^2 \right)^{\frac{1}{2}}$$

which yields

$$(24) \quad \begin{aligned} h \|\pi_0 f + \Delta u_h - \tau \pi_0 u_h\|_K + \|\tilde{h}^{\frac{1}{2}} [\nabla u_h]\|_{\mathcal{F}_i} + \|\tilde{h}^{-\frac{1}{2}} \overline{[u_h]}\|_{\mathcal{F}} \\ \leq ch (\|f - \pi_0 f\|_K + \tau h \|f\|_K). \end{aligned}$$

Finally observe that

$$\|f + \Delta u_h - \tau u_h\|_{\mathcal{K}} \leq \|f - \pi_0 f\|_{\mathcal{K}} + \tau \|u_h - \pi_0 u_h\|_{\mathcal{K}} + \|\pi_0 f + \Delta u_h - \tau \pi_0 u_h\|_{\mathcal{K}}$$

which, combined with (23) and (24), leads to (21) and using that  $\|f - \pi_0 f\|_{\mathcal{K}} \leq ch \|\nabla f\|_{\mathcal{K}}$  for  $f \in H^1(\mathcal{K})$  we prove (22).  $\square$

By choosing  $f \in V_h^0$  and  $\tau = 0$  in Proposition 4.10 above we immediately obtain the following Corollary.

**Corollary 4.11.** *If  $\tau = 0$  and  $f$  is piecewise constant, i.e.  $f \in V_h^0$ , then*

$$\|\overline{[u_h]}\|_{\mathcal{F}} = 0, \quad \|\nabla u_h\|_{\mathcal{F}_i} = 0 \quad \text{and} \quad \|\Delta u - \Delta u_h\|_{\mathcal{K}} = 0.$$

**4.4. Relation between the symmetric BSDG-method, the non-symmetric BSDG-method and mixed methods.** In this section we will explore some of the consequences of Corollary 4.11. To this end we assume that the function  $f$  is piecewise constant per element and that  $\tau = 0$ . Under this assumption we show that the solution of the symmetric and the non-symmetric discontinuous Galerkin methods are identical. Moreover we show that the method can be written as a method using Lagrange multipliers to impose continuity. Finally we consider the classical mixed method and show that the solution to the BSDG-method is equivalent to the solution of the mixed method.

Let us first define another useful projection.

**Lemma 4.12.** *Let  $a_h \in V_h^0$  be fixed. Then, there exists a unique function  $\phi_h \in V_{bs}$  such that*

$$(25) \quad \begin{cases} \Delta \phi_h &= a_h, \\ \overline{[\phi_h]_s}|_F &= 0 \quad \forall F \in \mathcal{F}, \\ [\nabla \phi_h]|_F &= 0 \quad \forall F \in \mathcal{F}_i. \end{cases}$$

Moreover  $\phi_h$  satisfies the following stability result

$$(26) \quad \|\phi_h\|^2 \leq c \|\tilde{h}^{-1} a_h\|_{\mathcal{K}}^2.$$

*Proof.* We proceed similarly as in the case of Lemma 3.8 and 3.9. Details are left to the reader.  $\square$

**Corollary 4.13.** *If  $\tau = 0$  and  $f$  is piecewise constant, i.e.  $f \in V_h^0$ , then the solution  $u_h$  of (12) is equal to the unique solution of the projection defined by Lemma 4.12 with  $a_h = f$ .*

**4.4.1. The symmetric and non-symmetric version have the same solutions.** Consider equation (11) with  $\tau = 0$  and its discretization using a non-symmetric scheme defined by: find  $u_{h,ns} \in V_{bs}$  such that

$$(27) \quad a_{ns}(u_{h,ns}, v_h) = (f, v_h)_{\mathcal{K}} \quad \forall v_h \in V_{bs},$$

with

$$a_{ns}(u_{h,ns}, v_h) = (\nabla u_{h,ns}, \nabla v_h)_{\mathcal{K}} - (\{\nabla u_{h,ns}\}, [v_h])_{\mathcal{F}} + ([u_{h,ns}], \{\nabla v_h\})_{\mathcal{F}}.$$

**Lemma 4.14.** *Let  $u_{h,ns}$  be the solution of (27) and  $u_h$  the solution of (12) with  $\tau = 0$ . If  $f \in V_h^0$ , then*

$$u_{h,ns} = u_h.$$

*Proof.* Using the same arguments as in Proposition 4.10 but for the non-symmetric scheme it can be shown that

$$\begin{aligned}\Delta u_{h,ns} &= f, \\ \overline{[u_{h,ns}]_s}|_F &= 0 \quad \forall F \in \mathcal{F}, \\ [\nabla u_{h,ns}]|_F &= 0 \quad \forall F \in \mathcal{F}_i.\end{aligned}$$

Thus  $u_{h,ns}$  is the unique projection defined by Lemma 4.12 with  $a_h = f$  and thus by Corollary 4.13 and the uniqueness of the projection we conclude that  $u_{h,ns} = u_h$ .  $\square$

*Remark 4.15.* In case  $f \in H^1(\mathcal{K})$  we may easily use the same argument to show that

$$\|u_h - u_{h,ns}\| \leq ch^2 \|\nabla f\|_{\mathcal{K}},$$

which in its turn implies that the error of the non-symmetric approximation  $u_{h,ns}$  enjoys optimal convergence in the  $L^2$ -norm when  $f \in H^1(\mathcal{K})$ .

**4.4.2. BSDG is equivalent to imposing continuity by Lagrange multipliers.** Consider again equation (11) with  $\tau = 0$ . Then, introduce the following Lagrange multiplier method by: find  $u_{h,L} \in V_{bs}$ ,  $\lambda_h \in W_h^0$  such that

$$(28) \quad \begin{cases} (\nabla u_{h,L}, \nabla v_h)_{\mathcal{K}} - (\lambda_h, [v_h]_s)_{\mathcal{F}} &= (f, v_h)_{\mathcal{K}} \quad \forall v_h \in V_{bs}, \\ ([u_{h,L}]_s, \mu_h)_{\mathcal{F}} &= 0 \quad \forall \mu_h \in W_h^0. \end{cases}$$

**Lemma 4.16.** *Let  $u_{h,L}$ ,  $\lambda_h$  be the solution of (28) and  $u_h$  the solution of (12). If  $f \in V_h^0$ , then*

$$\lambda_h|_F = \{\nabla u_h\}_s|_F \quad \forall F \in \mathcal{F} \quad \text{and} \quad u_{h,L} = u_h.$$

*Proof.* Observe that choosing  $\mu_h = \overline{[u_{h,L}]_s}$  in the second equation of (28) yields  $\|[\overline{u_{h,L}}]_s\|_{\mathcal{F}} = 0$ . Further let  $\phi_h \in V_{bs}$  be the projection defined by Lemma 3.8 with  $a_h = f + \Delta u_{h,L}$ ,  $b_h = \lambda_h - \{\nabla u_{h,L}\}_s$ ,  $c_h = -[\nabla u_{h,L}]$ . Then, choosing  $v_h = \phi_h$  in the first equation of (28) yields

$$\begin{aligned}0 &= (f, \phi_h)_{\mathcal{K}} - (\nabla u_{h,L}, \nabla \phi_h)_{\mathcal{K}} + (\lambda_h, [\phi_h]_s)_{\mathcal{F}} \\ &= (f + \Delta u_{h,L}, \pi_0 \phi_h)_{\mathcal{K}} - ([\nabla u_{h,L}], \{\overline{\phi_h}\})_{\mathcal{F}_i} + (\lambda_h - \{\nabla u_{h,L}\}_s, \overline{[\phi_h]_s})_{\mathcal{F}} \\ &= \|f + \Delta u_{h,L}\|_{\mathcal{K}}^2 + \|[\nabla u_{h,L}]\|_{\mathcal{F}_i}^2 + \|\lambda_h - \{\nabla u_{h,L}\}_s\|_{\mathcal{F}}^2.\end{aligned}$$

Thus we conclude firstly that

$$(29) \quad \{\nabla u_{h,L}\}_s|_F = \lambda_h|_F, \quad \forall F \in \mathcal{F}$$

and secondly that  $u_{h,L}$  is the unique projection defined by Lemma 4.12 with  $a_h = f$  and thus by corollary 4.13 and the uniqueness of the projection we conclude that  $u_{h,L} = u_h$ .  $\square$

**4.4.3. BSDG is equivalent to the classical mixed method.** We consider the standard mixed formulation using the lowest order Raviart-Thomas space for the fluxes and piecewise constants for the primal variable. The discretization of the elliptic problem (11) with  $\tau = 0$  takes the form: Find  $(w_h, u_h^0) \in RT_0 \times V_h^0$  such that

$$(30) \quad \begin{aligned}-(\nabla \cdot w_h, q_h)_{\mathcal{K}} &= (f, q_h)_{\mathcal{K}}, \\ (w_h, v_h)_{\mathcal{K}} + (u_h^0, \nabla \cdot v_h)_{\mathcal{K}} &= 0\end{aligned}$$

for all  $(v_h, q_h) \in RT_0 \times V_h^0$ . Noting now that, using the solution  $(w_h, u_h^0)$ , a reconstruction of the primal variable can be made in the space  $V_{bs}$ . We denote

this reconstruction  $\nu_h = u_h^0 + \tilde{u}_h$  where  $\tilde{u}_h \in V_{bs}$  is the solution to the problem  $\nabla \tilde{u}_h|_\kappa = w_h|_\kappa$  and  $\int_\kappa \tilde{u}_h = 0$  for all  $\kappa \in \mathcal{K}$ . Lemma 3.2 proves uniqueness of  $\tilde{u}_h|_\kappa$  under the assumption of  $\int_\kappa \tilde{u}_h = 0$ . Using the reconstruction we introduce the following box-scheme: Find  $(w_h, u_h^0) \in RT_0 \times V_h^0$  such that

$$(31) \quad \begin{aligned} -(\nabla \cdot w_h, q_h)_\mathcal{K} &= (f, q_h)_\mathcal{K}, \\ ([u_h^0 + \tilde{u}_h], v_h)_\mathcal{F} &= 0 \end{aligned}$$

for all  $(v_h, q_h) \in RT_0 \times V_h^0$ .

**Lemma 4.17.** *The scheme defined by (31) has a unique solution.*

*Proof.* This formulation results in a square matrix since the number of unknowns and the number of equations are the same. Existence and uniqueness of a discrete solution is easily shown assuming  $f = 0$  and proving that  $w_h = 0, u_h^0 = 0$ . By integration by parts we have using the second equation of (31) and the zero average property of  $\tilde{u}_h$ ,

$$\|w_h\|_\mathcal{K}^2 = \|\nabla \tilde{u}_h\|_\mathcal{K}^2 = (\nabla(u_h^0 + \tilde{u}_h), \nabla \tilde{u}_h)_\mathcal{K} = -\underbrace{(\Delta \tilde{u}_h, \tilde{u}_h)_\mathcal{K}}_{=0} + \underbrace{([u_h^0 + \tilde{u}_h], \nabla \tilde{u}_h)_\mathcal{F}}_{=0} = 0.$$

The second equation of (31) now immediately gives  $u_h^0 = 0$  when testing with  $v_h$  such that  $v_h|_F = [u_h^0]|_F$ .  $\square$   $\square$

**Lemma 4.18.** *The solution of (31) is equivalent to the solution of (30).*

*Proof.* Since the first equation is the same in the two cases, it remains to show that the second equation of the two problems (30) and (31) are equivalent. Note that by definition

$$(w_h, v_h)_\mathcal{K} + (u_h^0, \nabla \cdot v_h)_\mathcal{K} = (\nabla \tilde{u}_h, v_h)_\mathcal{K} + (u_h^0, \nabla \cdot v_h)_\mathcal{K}.$$

Integrating by parts in both terms of the right hand side and using the zero mean property of  $\tilde{u}_h$  we obtain

$$(w_h, v_h)_\mathcal{K} + (u_h^0, \nabla \cdot v_h)_\mathcal{K} = ([u_h^0 + \tilde{u}_h], v_h)_\mathcal{F} - \underbrace{(\tilde{u}_h, \nabla \cdot v_h)_\mathcal{K}}_{=0} = ([u_h^0 + \tilde{u}_h], v_h)_\mathcal{F}.$$

$\square$

$\square$

**Lemma 4.19.** *If  $f \in V_h^0$ , then the reconstructed solution  $\nu_h = u_h^0 + \tilde{u}_h$  of (31) (and (30)) is also solution to (12), i.e.  $\nu_h = u_h$  where  $u_h$  is the solution of (12).*

*Proof.* Observe by the first equation of (31) that

$$-(\Delta \nu_h, q_h)_\mathcal{K} = -(\Delta \tilde{u}_h, q_h)_\mathcal{K} = -(\nabla \cdot w_h, q_h)_\mathcal{K} = (f, q_h)_\mathcal{K}$$

for all  $q_h \in V_h^0$  and thus  $-\Delta \nu_h = f$ . Further there exists a function  $\tilde{v}_h \in RT_0$  such that  $\tilde{v}_h|_F = [\nu_h]|_F$  for all  $F \in \mathcal{F}$ . Thus by the second equation of (31) with  $v_h = \tilde{v}_h$  we get  $\|[\nu_h]\|_\mathcal{F} = \|[\nu_h]_s\|_\mathcal{F} = 0$ . Observe that  $\nabla \nu_h = w_h \in RT_0$  and therefore  $\|[\nabla \nu_h]\|_{\mathcal{F}_i} = 0$ . Again we conclude by noting that  $\nu_h$  is the unique projection defined by Lemma 4.12 with  $a_h = f$  and thus by corollary 4.13 and the uniqueness of the projection we conclude that  $\nu_h = u_h$ .  $\square$   $\square$

## 5. PARABOLIC PROBLEMS

Let us consider the following parabolic problem: find  $u : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}$  such that

$$(32) \quad \begin{cases} \partial_t u - \Delta u &= f & \text{in } \Omega \times \mathbb{R}^+, \\ u &= 0 & \text{on } \partial\Omega \times \mathbb{R}^+, \\ u|_{t=0} &= u_0 & \text{on } \Omega, \end{cases}$$

with  $\partial_t u \equiv \frac{\partial u}{\partial t}$ ,  $f \in L^2(\Omega \times \mathbb{R}^+)$  and  $u_0 \in L^2(\Omega)$ . Then using the discretization technique of Section 4 the discretization in space of (32) reads: For all  $t > 0$  find  $u_h(t) \in V_{bs}$  such that

$$(33) \quad \begin{cases} (\partial_t u_h(t), v_h)_K + a(u_h(t), v_h) &= (f(t), v_h)_K \quad \forall v_h \in V_{bs}, \\ u_h(0) &= u_{h,0}, \end{cases}$$

where the bilinear form  $a(\cdot, \cdot)$  is defined by (14),  $f \in L^2(\Omega \times \mathbb{R}^+)$  and  $u_{h,0} \in V_{bs}$  is some approximation of  $u_0$  in  $V_{bs}$  satisfying

$$(34) \quad \|u_{h,0} - u_0\|_K + h\|u_{h,0} - u_0\| \leq c h^2 |u_0|_{2,K}.$$

Let  $\delta t > 0$  be the time step,  $u_h^n$  the approximation of  $u$  at time  $t_n = n \delta t$  and denote by  $\bar{\partial}_t u_h^n$  the backward difference quotient defined by  $\bar{\partial}_t u_h^n = (u_h^n - u_h^{n-1})/\delta t$ . Then, the backward Euler time-discretization reads: find  $\{u_h^n\}_{n \geq 1} \subset V_{bs}$  such that

$$(35) \quad \begin{cases} (\bar{\partial}_t u_h^n, v_h)_K + a(u_h^n, v_h) &= (f^n, v_h)_K \quad \forall v_h \in V_{bs}, \\ u_h^0 &= u_{h,0}, \end{cases}$$

where we denote by  $f^n = f(t_n)$ .

**Lemma 5.1** (Consistency). *Let  $u(t_n) \in H^2(\Omega)$  be the exact solution of (32) at time  $t_n$  and let  $u_h^n \in V_{bs}$  be the approximation defined by (35), then the method satisfies the following Galerkin orthogonality*

$$(\partial_t u(t_n) - \bar{\partial}_t u_h^n, v_h)_K + a(u(t_n) - u_h^n, v_h) = 0 \quad \forall v_h \in V_{bs}.$$

*Proof.* The technical details of this proof are identical with the ones of the proof of Lemma 4.2.  $\square$   $\square$

**5.1. Stability.** In (13) and (15) of the definitions of the bilinear form  $A(\cdot, \cdot)$  and the energy-norm  $\|\cdot\|_e$  we choose now  $\tau = 1/\delta t$  which yields, for parabolic problems,

$$\begin{aligned} A(v, w) &= a(v, w) + \frac{1}{\delta t} (v, w)_K, \\ \|v\|_e^2 &= \frac{c_d}{4} \|v\|^2 + \|\delta t^{-\frac{1}{2}} v\|_K^2, \end{aligned}$$

where the bilinear form  $a(\cdot, \cdot)$  is still the one defined by (14) and  $c_d > 0$  the constant defined by (3).

**Proposition 5.2** (Stability). *If  $h^2 < c_s \delta t$  for some constant  $c_s > 0$  independent of  $h$ ,  $N$  and  $\delta t$ , then the approximation of (35) satisfies the following stability result*

$$\|u_h^N\|_K \leq \|u_{h,0}\|_K + \delta t \sum_{n=1}^N \|f^n\|_K.$$

*Proof.* Let  $w_h^n = w_h^n(u_h^n) = u_h^n + c \phi_h^n$  be the function defined by Lemma 4.4. Remember that  $\phi_h^n$  is the projection defined by Lemma 3.9 with  $a_h = 0$ ,  $b_h = -\tilde{h}^{-1} \overline{[u_h^n]_s}$  and  $c_h = 0$ . Then, choose  $v_h = w_h^n$  in (35) and thus

$$(36) \quad A(u_h^n, w_h^n) = (f^n, w_h^n)_K + \frac{1}{\delta t} (u_h^{n-1}, w_h^n)_K.$$

Observe that by Lemma 4.4 we get

$$(37) \quad |||u_h^n|||_e^2 \leq A(u_h^n, w_h^n).$$

Applying the Cauchy-Schwarz inequality and the stability estimate (10) further yields

$$|||w_h^n|||_{\mathcal{K}}^2 \leq |||u_h^n|||_{\mathcal{K}}^2 + c||\phi_h^n||_{\mathcal{K}}^2 \leq |||u_h^n|||_{\mathcal{K}}^2 + ch^2|||u_h^n|||_{\mathcal{K}}^2 \leq |||u_h^n|||_{\mathcal{K}}^2 + \frac{\delta t}{4}|||u_h^n|||_{\mathcal{K}}^2 \leq \delta t|||u_h^n|||_e^2.$$

Here we used the assumption that  $h^2/\delta t$  is small enough. Next we develop

$$(38) \quad (f^n, w_h^n)_{\mathcal{K}} \leq \|f^n\|_{\mathcal{K}} |||w_h^n|||_{\mathcal{K}} \leq \delta t^{\frac{1}{2}} \|f^n\|_{\mathcal{K}} |||u_h^n|||_e$$

and

$$(39) \quad \frac{1}{\delta t}(u_h^{n-1}, w_h^n)_{\mathcal{K}} \leq \frac{1}{\delta t}|||u_h^{n-1}|||_{\mathcal{K}} |||w_h^n|||_{\mathcal{K}} \leq \frac{1}{\delta t^{\frac{1}{2}}}|||u_h^{n-1}|||_{\mathcal{K}} |||u_h^n|||_e.$$

Respecting the three bounds (37)-(39) in (36) yields

$$|||u_h^n|||_e \leq \delta t^{\frac{1}{2}} \|f^n\|_{\mathcal{K}} + \delta t^{-\frac{1}{2}} |||u_h^{n-1}|||_{\mathcal{K}}$$

and thus

$$|||u_h^n|||_{\mathcal{K}} \leq \delta t \|f^n\|_{\mathcal{K}} + |||u_h^{n-1}|||_{\mathcal{K}}$$

by definition of the norm  $||| \cdot |||_e$ . Summing the previous expression from  $n = 1$  to  $n = N$  leads to

$$|||u_h^N|||_{\mathcal{K}} \leq |||u_{h,0}|||_{\mathcal{K}} + \delta t \sum_{n=1}^N \|f^n\|_{\mathcal{K}}.$$

□

□

**5.2. Convergence analysis.** Let us introduce the Ritz-projection associated to the BSDG-formulation. In order to have optimal approximation properties also in the  $L^2$ -norm we assume from now on that  $\Omega$  is convex. Let  $v \in H^2(\Omega)$  be fixed and consider the following projection: find  $P_h v \in V_{bs}$  such that

$$(40) \quad a(P_h v, v_h) = a(v, v_h), \quad \forall v_h \in V_{bs}.$$

This problem formulation corresponds to the one of (12) with  $\tau = 0$  and the particular choice of the right hand side  $a(v, \cdot)$ . Thus, the projection exists and satisfies the following approximation result

$$(41) \quad \|P_h v - v\|_{\mathcal{K}} + h|||P_h v - v||| \leq ch^2|v|_{2,\mathcal{K}}$$

according to Theorem 4.8 and 4.9.

**Theorem 5.3** (Convergence). *Let  $u$  be the exact solution of problem (32) and let  $u_h^N$  be the approximation defined by (35) at time  $t_N$ . If  $h^2 < c_s \delta t$  for some constant  $c_s > 0$  independent of  $h$ ,  $N$  and  $\delta t$ , then there holds*

$$|||u_h^N - u(t_N)|||_{\mathcal{K}} \leq ch^2 \left( |u_0|_{2,\mathcal{K}} + \int_0^{t_N} |u_t(s)|_{2,\mathcal{K}} ds \right) + c \delta t \left( \int_0^{t_N} |||u_{tt}(s)|||_{\mathcal{K}} ds \right)$$

where the constant  $c > 0$  is independent of  $h$ ,  $\delta t$  and  $N$ .

*Proof.* This proof is similar to the proof of Theorem 1.5 in [13], but applied to our method. Firstly split the error into two parts

$$u_h^N - u(t_N) = u_h^N - P_h u(t_N) + P_h u(t_N) - u(t_N) = \xi_h^N + \eta(t_N)$$

with  $\xi_h^N = u_h^N - P_h u(t_N)$ ,  $\eta(t_N) = P_h u(t_N) - u(t_N)$  and apply the triangle inequality

$$(42) \quad |||u_h^N - u(t_N)|||_{\mathcal{K}} \leq |||\xi_h^N|||_{\mathcal{K}} + |||\eta(t_N)|||_{\mathcal{K}}.$$

Observe that the second term of the right hand side of (42) is bounded by

$$(43) \quad \|\eta(t_N)\|_{\mathcal{K}} \leq ch^2 |u(t_N)|_{2,\mathcal{K}} \leq ch^2 \left( |u_0|_{2,\mathcal{K}} + \int_0^{t_N} |u_t(s)|_{2,\mathcal{K}} ds \right).$$

On the other hand, using the same notation as in the previous section, i.e.  $\bar{\partial}_t \xi_h^n = (\xi_h^n - \xi_h^{n-1})/\delta t$ , observe that applying the consistency of the method, Lemma 5.1, and the definition of the projection  $P_h$  defined in (40) yields

$$\begin{aligned} & (\bar{\partial}_t \xi_h^n, v_h)_{\mathcal{K}} + a(\xi_h^n, v_h) \\ &= (\bar{\partial}_t u_h^n, v_h)_{\mathcal{K}} - (\bar{\partial}_t P_h u(t_n), v_h)_{\mathcal{K}} + a(u_h^n - P_h u(t_n), v_h) \\ &= (\bar{\partial}_t u(t_n), v_h)_{\mathcal{K}} - (\bar{\partial}_t P_h u(t_n), v_h)_{\mathcal{K}} + a(u(t_n) - P_h u(t_n), v_h) \\ &= (\partial_t u(t_n) - \bar{\partial}_t P_h u(t_n), v_h)_{\mathcal{K}} = (w(t_n), v_h)_{\mathcal{K}} \end{aligned}$$

for all  $v_h \in V_{bs}$  and with  $w(t_n) = \partial_t u(t_n) - \bar{\partial}_t P_h u(t_n)$ . Remark that the function  $w$  can alternatively be defined as  $w(t_n) = \partial_t u(t_n) - P_h \partial_t u(t_n)$  by linearity of  $P_h$ . Thus  $\{\xi_h^n\}_{n=0}^N$  satisfies the scheme (35) with particular choice of  $f = w$  and therefore the following stability holds

$$(44) \quad \|\xi_h^N\|_{\mathcal{K}} \leq \|\xi_h^0\|_{\mathcal{K}} + \delta t \sum_{n=1}^N \|w(t_n)\|_{\mathcal{K}}$$

by Proposition 5.2. Using (34) we get

$$(45) \quad \|\xi_h^0\|_{\mathcal{K}} \leq ch^2 |u_0|_{2,\mathcal{K}}.$$

The second term of the right hand side of (44) is bounded exactly as in the proof of Theorem 1.5 in [13].  $\square$   $\square$

**5.3. Numerical tests.** Let us introduce the problems we used for the numerical tests:

i) *Problem with smooth solution*

Let  $d = 2, 3$  denote the dimension and let  $\Omega = (0, 1)^d$ . Further denote by  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  the spatial variable. Consider problem (32) with  $f(x, t) = -(d\pi^2 + 0.1)e^{-0.1t} \sin(\pi x_1) \cos(\pi x_2) \cos(\pi x_3)$ , non-homogeneous boundary conditions and initial condition such that the solution is  $u(x, t) = e^{-0.1t} \sin(\pi x_1) \cos(\pi x_2) \cos(\pi x_3) \in C^\infty(\Omega)$ . Assume that in the two-dimensional case there holds  $x_3 = 0$ .

ii) *Problem with non-smooth solution*

Now choose the following  $L$ -shaped domain:  $\Omega = ([-1, 1] \times [-1, 0] \cup [0, 1]^2)^\circ \subset \mathbb{R}^2$ . We consider problem (32) with

$$f(x, t) = -0.1e^{-0.1t} (x_1^2 + x_2^2)^{\frac{1}{3}} \sin\left(\frac{2}{3} \arctan_*\left(\frac{x_1}{x_2}\right)\right),$$

non-homogeneous boundary conditions and initial condition such that the solution is

$$u(x, t) = e^{-0.1t} (x_1^2 + x_2^2)^{\frac{1}{3}} \sin\left(\frac{2}{3} \arctan_*\left(\frac{x_1}{x_2}\right)\right)$$

where  $\arctan_*$  is chosen in the manner that it is a continuous function at points with  $x_2 = 0$ . One can prove that  $u \notin H^2(\Omega)$ . Therefore the theory presented in this paper is no longer valid.



$h$	$\delta t$	$\ u(t_N) - u_h^N\ $	$\ u(t_N) - u_h^N\ _{\mathcal{K}}$	$\ \Delta(u(t_N) - u_h^N)\ _{\mathcal{K}}$	$\ \tilde{h}^{\frac{1}{2}}[\nabla u_h^N]\ _{\mathcal{F}_i}$
0.2	1	3.88E-01	1.84E-02	1.89	1.33E-01
0.1	0.25	1.94E-01 (1.00)	4.69E-03 (1.97)	8.86E-01 (1.09)	2.57E-02 (2.37)
0.05	0.0625	1.00E-01 (0.95)	1.26E-03 (1.90)	4.61E-01 (0.94)	6.74E-03 (1.93)
0.025	0.015625	5.10E-02 (0.97)	3.27E-04 (1.94)	2.40E-01 (0.94)	1.91E-03 (1.82)

TABLE 1. *Different error-quantities at time  $t_N = 1$  for the problem i) with smooth solution in two spatial dimensions. The value in the bracket corresponds to the convergence rate.*

$h$	$\delta t$	$\ u(t_N) - u_h^N\ $	$\ u(t_N) - u_h^N\ _{\mathcal{K}}$	$\ \Delta(u(t_N) - u_h^N)\ _{\mathcal{K}}$	$\ \tilde{h}^{\frac{1}{2}}[\nabla u_h^N]\ _{\mathcal{F}_i}$
0.2	1	1.34E-01	1.13E-02	7.54E-03	5.94E-04
0.1	0.25	8.89E-02 (0.60)	4.86E-03 (1.21)	1.22E-03 (2.63)	1.09E-05 (5.78)
0.05	0.0625	5.78E-02 (0.62)	1.96E-03 (1.31)	3.96E-04 (1.62)	1.97E-06 (2.46)
0.025	0.015625	3.62E-02 (0.67)	7.49E-04 (1.39)	1.20E-04 (1.73)	3.06E-07 (2.69)

TABLE 2. *Different error-quantities at time  $t_N = 1$  for the problem ii) with non-smooth solution in two spatial dimensions. The value in the bracket corresponds to the convergence rate.*

In both cases a sequence of unstructured meshes is considered. For the implementation of the bubble enriched finite element space  $V_{bs}$  observe that we can write

$$V_{bs} = V_h^1 \oplus b(x) V_h^0$$

where the quadratic function  $b$  is defined by  $b(x) = x \cdot x$ . Thus, for example equation (12) can be written as: find  $u_{h,1} \in V_h^1$  and  $u_{h,0} \in V_h^0$  such that

$$\begin{cases} A(u_{h,1}, v_{h,1}) + A(bu_{h,0}, v_{h,1}) &= (f, v_{h,1})_{\mathcal{K}} \quad \forall v_{h,1} \in V_h^1, \\ A(u_{h,1}, bv_{h,0}) + A(bu_{h,0}, bv_{h,0}) &= (f, bv_{h,0})_{\mathcal{K}} \quad \forall v_{h,0} \in V_h^0. \end{cases}$$

This splitting is also valid for the implementation of the scheme for parabolic problems of the form (35) in a similar manner. Using this splitting the bubble stabilized discontinuous Galerkin method can be implemented using a finite element code that handles discontinuous approximation spaces and mixed formulations. For the computations the C++ library *life* is used, see [10, 11].

**5.3.1. Smooth solution.** Consider the numerical scheme (35) for the test problem i) in two and three spatial dimensions. Table 1 shows the convergence behavior of the error measured at time  $t_N = 1$  for different values of  $h$  and  $\delta t$  for the 2D-problem. Observe that  $\delta t$  is proportional to  $h^2$  according to Theorem 5.3. Additionally Figure 1 illustrates the convergence behavior of the energy- and  $L^2$ -norm for the same example but also for the three dimensional problem. We observe optimal convergence as predicted by the theory.

**5.3.2. Non-smooth solution.** Table 2 illustrates the convergence behavior of the approximation for the problem ii) with non-smooth solution. Note the super-convergence of  $\|\Delta u(t_N) - \Delta u_h^N\|_{\mathcal{K}}$ .

**5.3.3. Second order time solver.** Instead of using the backward Euler scheme to solve the time dependence we test here the Crank-Nicolson scheme. Thus, we use also a second order accurate method with respect to the time step. Table 3 shows the convergence behavior of the resulting approximation. Observe that the

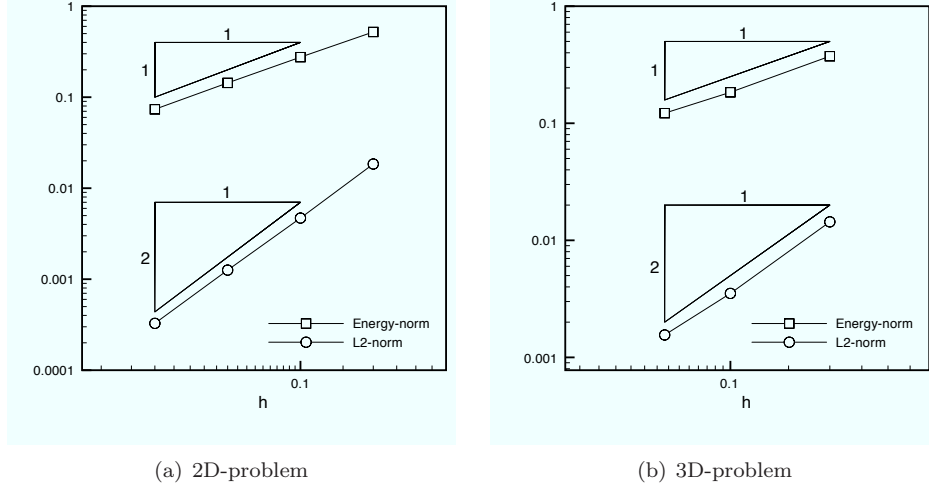


FIGURE 1. Two- and three-dimensional computations for the test problem i) with smooth solution. The error is measured at time  $t_N = 1$  and  $\delta t$  is proportional to  $h^2$ .

$h$	$\delta t$	$\ u(t_N) - u_h^N\ $	$\ u(t_N) - u_h^N\ _{\mathcal{K}}$	$\ \Delta(u(t_N) - u_h^N)\ _{\mathcal{K}}$	$\ \bar{h}^{\frac{1}{2}} [\nabla u_h^N]\ _{\mathcal{F}_i}$
0.1	0.1	1.39E-01	2.10E-03	2.75	4.66E-01
0.05	0.05	7.35E-02 (0.92)	5.76E-04 (1.87)	3.05 (-0.15)	2.52E-01 (0.88)
0.025	0.025	3.71E-02 (0.99)	1.47E-04 (1.97)	3.11 (-0.03)	1.27E-01 (0.99)
0.0125	0.0125	1.83E-02 (1.02)	3.70E-05 (1.99)	2.96 (0.07)	6.23E-02 (1.03)

TABLE 3. Different error-quantities at time  $t_N = 1$  for the problem i) with smooth solution in two spatial dimensions using the Crank-Nicolson formula for solving the time-dependency. The value in the bracket corresponds to the convergence rate.

method is now second order accurate in the  $L^2$ -norm with respect to the spatial and temporal discretization parameters, i.e.  $\delta t$  is now chosen proportionally to  $h$ . On the other hand observe that the convergence of  $\|\Delta(u(t_N) - u_h^N)\|_{\mathcal{K}}$  is destroyed by applying the Crank-Nicolson formula.

**5.3.4. Stability.** We address here the stability of the numerical scheme. Remember that the stability result of Proposition 5.2 is obtained under the condition that  $h^2/\delta t$  is small enough, i.e.  $h^2 \leq c_s \delta t$  for some constant  $c_s$ . We will quantify this constant numerically in the case of the test problems i) and ii) in two dimensions. Observe that the scheme (35) has the following form in its matrix-vector formulation

$$(M + \delta t A) \mathbf{u}^n = M \mathbf{u}^{n-1} + \delta t \mathbf{f}^n$$

where  $M$  is the mass matrix,  $A$  the matrix corresponding to the bilinear form  $a(\cdot, \cdot)$ ,  $\mathbf{u}^n$  the solution vector at time  $t_n$  and  $\mathbf{f}^n$  the vector associated to the right hand side at time  $t_n$ . Thus the scheme is stable if and only if the spectral radius  $\rho$  of the

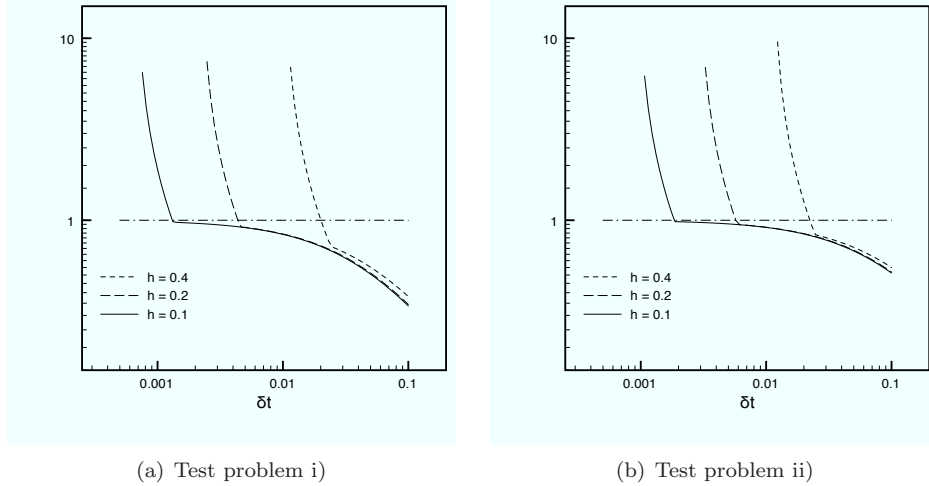


FIGURE 2. The spectral radius of the iteration matrix of the numerical scheme (35) for different mesh sizes  $h$  and depending on the temporal discretization step  $\delta t$  for the test problems i) and ii).

$h$	$\delta t^*$	$h^2/\delta t^*$
0.4	0.0215	7.442
0.2	0.0046	8.696
0.1	0.00132	7.576

$h$	$\delta t^*$	$h^2/\delta t^*$
0.4	0.0231	6.926
0.2	0.00572	6.993
0.1	0.00187	5.337

TABLE 4. Quantification of the stability constant for the test problem i) and ii). The quantity  $\delta t^*$  denotes the smallest numerically tested value of  $\delta t$  such that  $\rho(B(\delta t)) < 1$ .

iteration matrix  $B(\delta t) = (M + \delta t A)^{-1} M$  is smaller than 1, i.e.

$$\rho(B(\delta t)) = \rho((M + \delta t A)^{-1} M) < 1.$$

Figure 2 shows the spectral radius of the iteration matrix  $B(\delta t) = (M + \delta t A)^{-1} M$  for different mesh sizes depending on the temporal discretization step  $\delta t$ . In Table 4 we illustrate the constant of the inequality  $h^2 \leq c_s \delta t$  for each  $h$ . Note that we use unstructured meshes.

#### ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided through the Swiss National Science Foundation under grant 200021 – 113304.

#### REFERENCES

- [1] T. Arbogast and Z. Chen. On the implementation of mixed methods as nonconforming methods for second-order elliptic problems. *Math. Comp.*, 64(211):943–972, 1995.
- [2] D. N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19(4):742–760, 1982.
- [3] G. A. Baker. Finite element methods for elliptic equations using nonconforming elements. *Math. Comp.*, 31(137):45–59, 1977.

- [4] F. Brezzi and L. D. Marini. Bubble stabilization of discontinuous Galerkin methods. In W. Fitzgibbon, R. Hoppe, J. Periaux, O. Pironneau, and Y. Vassilevski, editors, *Advances in numerical mathematics, Proc. International Conference on the occasion of the 60th birthday of Y.A. Kuznetsov*, pages 25–36. Institute of Numerical Mathematics of The Russian Academy of Sciences, Moscow, 2006.
- [5] E. Burman and B. Stamm. Low order discontinuous Galerkin methods for second order elliptic problems. Technical report, EPFL-IACS report 04, 2007. submitted to SIAM J. Numer. Anal.
- [6] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 40 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)].
- [7] J. Douglas, Jr. and T. Dupont. Interior penalty procedures for elliptic and parabolic Galerkin methods. In *Computing methods in applied sciences (Second Internat. Sympos., Versailles, 1975)*, pages 207–216. Lecture Notes in Phys., Vol. 58. Springer, Berlin, 1976.
- [8] M. G. Larson and A. J. Niklasson. Analysis of a nonsymmetric discontinuous Galerkin method for elliptic problems: stability and energy error estimates. *SIAM J. Numer. Anal.*, 42(1):252–264 (electronic), 2004.
- [9] J. T. Oden, I. Babuska, and C. E. Baumann. A discontinuous *hp* finite element method for diffusion problems. *J. Comput. Phys.*, 146(2):491–519, 1998.
- [10] C. Prud’homme. A domain specific embedded language in C++ for automatic differentiation, projection, integration and variational formulations. *Scientific Programming*, 14(2):81–110, 2006.
- [11] C. Prud’homme. A domain specific embedded language in C++ for automatic differentiation, projection, integration and variational formulations. *Scientific Programming*, 14(2):81–110, 2006.
- [12] B. Rivière, M. F. Wheeler, and V. Girault. A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems. *SIAM J. Numer. Anal.*, 39(3):902–931 (electronic), 2001.
- [13] V. Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.
- [14] M. F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15(1):152–161, 1978.

ERIK BURMAN, DEPARTMENT OF MATHEMATICS, MANTELL BUILDING, UNIVERSITY OF SUSSEX,  
FALMER, BRIGHTON, BN1 9RF, UNITED KINGDOM  
*E-mail address*: `E.N.Burman@sussex.ac.uk`

BENJAMIN STAMM, EPFL, IACS-CMCS, STATION 8, 1015 LAUSANNE, SWITZERLAND  
*E-mail address*: `benjamin.stamm@epfl.ch`